Implementing Bignums in $\mu$Smalltalk

REVISED

Norman Ramsey

Fall 2018

1 Approach

The pair portion of the $\mu$Smalltalk assignment is to implement arbitrary-precision arithmetic (“bignums”). You’ll write a lot of methods. To help you organize them, I suggest which methods to implement in what order, and I sketch how implementations of some methods may depend on other methods.

The diagram below shows what the class hierarchy will look like once you finish. The unboxed classes are predefined $\mu$Smalltalk classes. The boxed classes are new classes you will write for this assignment. Each class that is followed by a number is from the Exercise with that number.

\[
\text{Object} \quad \text{Magnitude} \quad \text{Number} \quad \text{Natural} \quad \text{Fraction} \quad \text{Float} \quad \text{Integer} \quad \text{SmallInteger (original)} \quad \text{LargeInteger} \quad \text{SmallInteger (redefined)} \quad \text{LargePositiveInteger} \quad \text{LargeNegativeInteger} \]

1.1 Big picture, part I: Natural numbers

Here is the big picture of which parts of the system do what. First, here is how you replicate the work you did on the SML assignment:

- *Comparisons* are implemented in class $\text{Magnitude}$. Or rather, four of the six comparisons are implemented in $\text{Magnitude}$. The fundamental
comparisons = and < are subclass responsibilities (see the definition of class Magnitude around page 881). You have to implement = and < on class Natural. The other four (!=, >, <=, and >=) are inherited from Magnitude.

- Addition, subtraction, multiplication, and (short) division are all implemented on class Natural.
  - Good news: the protocol guarantees that the argument, not just the receiver, of each of these methods has class Natural. No double dispatch is required.
  - Bad news: Smalltalk uses objects to hide information. Just because you know an object’s class doesn’t mean you have access to its representation. You will need private methods to get access to the degree and digits of the argument. The book recommends methods degree and digit:

1.2 Big picture, part II: Large integers

As described on page 949, large integers are implemented using sign-magnitude representation. The sign is encoded in the integer's class: a large negative integer has class LargeNegativeInteger, and a large nonnegative integer has class LargePositiveInteger.¹ The magnitude, which has class Natural, is stored in an instance variable, which I’ve called magnitude.

Both LargeNegativeInteger and LargePositiveInteger inherit from LargeInteger. Class LargeInteger is an abstract class which can define the instance variable magnitude and which can also define some methods that are common to both positive and negative large integers:

- Methods = and < can be implemented by subtracting the argument from self and examining the result to see if it is zero (or negative).

- I recommend defining a private method isZero which delegates to the integer’s magnitude.² This method will help your code work correctly with both “positive” and “negative” zero.

¹In the perverse jargon of Smalltalk, zero is considered “positive” and positive numbers are considered “strictly positive.”

²“Delegation” is a term of art for an implementation technique in which a method is implemented by sending the same message to another object. In this case, I am recommending that you implement isZero on a large-integer object by sending the isZero message to the object that represents that large integer’s magnitude.
Most operations on large integers require double dispatch. For example, to add two large integers, you have to know the sign of both argument and receiver. A receiver knows its own sign; the sign of an argument is communicated by double dispatch. For example, method + on class LargePositiveInteger looks like this:

\[
(method + (anInteger)
  (addLargePositiveIntegerTo: anInteger self))
\]

The sign of the argument is encoded in the message name addLargePositiveIntegerTo:, and the implementation of the method addLargePositiveIntegerTo: depends on the class of the receiver:

- When a positive large integer receives addLargePositiveIntegerTo:, it knows that the sum of two positive integers is positive, and it sends withMagnitude: to class LargePositiveInteger with the sum of the magnitudes.

- When a negative large integer receives addLargePositiveIntegerTo:, it sends subtract:withDifference:ifNegative: to the argument's magnitude, with a success continuation that produces a large positive integer and a failure continuation that produces a large negative integer.

All this is achieved without ever interrogating the class of an object, which keeps the system “open”—any object that has the right protocol will work.

Double dispatch is also used to implement multiplication, which is a little easier, because there is less case analysis—the product of two magnitudes is always a (nonnegative) magnitude.

Always write as little double dispatch as possible. For example, implement the negated method without extra dispatch—just create a new number with the same magnitude as the receiver but the opposite sign. And you can implement subtraction without writing any new dispatch at all! The default implementation, which dispatches to class Number, works perfectly well with large integers.

If you feel yourself not quite certain about double dispatch, you can read more about it in the book section “Inspecting multiple representations the object-oriented way: magnitudes and numbers” on page 880.

Once you have large integers working, you have a system that exemplifies the expressive power of Smalltalk: arithmetic and relational operators are implemented by messages flying around and dispatching on methods.
of classes Magnitude, Number, LargeInteger, LargePositiveInteger, and LargeNegativeInteger. Your final step is “mixed arithmetic” with large and small integers.

1.3 Big picture, part III: Mixed arithmetic

Mixed arithmetic has two goals:

- Seamlessly allow arithmetic and relational operators on a mix of small and large integers.
- Extend small-integer arithmetic so that when an intermediate result doesn’t fit in a machine word, it automatically “fails over” to large-integer arithmetic.

The net result should be a system of arithmetic where you “pay as you go”: if you don’t need the features of large arithmetic, you’re not paying extra for them, but if you do need them, they are there automatically.

You implement mixed arithmetic by applying one technique you’ve applied before, plus two new ones.

- The technique you’ve applied before is double dispatch. For example, to add a small integer to a number, you’ll need a new method addSmallIntegerTo:, and method + on a small integer will dispatch to addSmallIntegerTo:. New method addSmallIntegerTo: must be defined on both large and small integers.

- The first new technique is coercion, which you can study in the context of classes Integer, Fraction, and Float. Whenever you perform an operation on mixed large and small integers, you coerce the small integer to a large one and repeat the operation. This form of coercion is the same regardless of the sign of the large integer, so it can go on class LargeInteger.

- The second new technique is to use primitives that detect overflow. For example, when adding small integers, you can no longer use primitive +, because it doesn’t handle overflow. You’ll need a different primitive that can invoke a failure continuation when addition overflows.
2 Natural numbers: Arrays or lists?

The book assumes that you will represent a sequence of digits using a Smalltalk array. While there is something to be said for following the advice in the book, there is also something to be said for building off of a list-based solution that is known to be good. If you prefer a list, be aware that the solution is still going to be quite a bit different than what you’ve seen in ML. My solutions to the ML modules homework demonstrate both list and array representations, so you have a choice of two equally good starting points. Here’s how I view the tradeoff:

- Using the list representation, the algorithms and the individual methods are simpler, but there’s a lot of dynamic dispatch, and if you want to understand the system, you’re going to have to develop a reasonably deep understanding of dynamic dispatch and object-oriented design.

- Using the array representation, there are more traps and pitfalls around the algorithms and mechanisms, but there’s not nearly as much dynamic dispatch—for big stretches of the work, you’ll be able to pretend that you’re programming in C or C++.

Which experience you want to have is up to you.

3 Details of class Natural, list version

If you want to try representing a natural number as a list of digits, I recommend that you avoid the existing List representation. Instead, define your own versions of empty and nonempty lists as subclasses of class Natural. An empty natural number represents zero and has no instance variables. A nonempty natural number has two instance variables: the least significant digit, which is a SmallInteger, and the remaining digits, which are an object of class Natural. I recommend one invariant, stated two ways:

- Zero is always represented as an object of the empty class.

- Every object of the nonempty class represents a natural number strictly greater than zero.

This invariant simplifies the implementations of many methods, including addition, subtraction, multiplication, division, comparison, and isZero. To maintain the invariant, you’ll have to implement the private class method
Private class methods for class `Natural`:

<table>
<thead>
<tr>
<th>Method</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><code>base</code></td>
<td>Answer the base of natural numbers.</td>
</tr>
<tr>
<td><code>first:rest: anInteger aNatural</code></td>
<td>Answers a <code>Natural</code> number representing $a\text{Natural} \cdot b + \text{anInteger}$, where $b$ is the base of <code>Natural</code> numbers.</td>
</tr>
<tr>
<td><code>oneDigit: anInteger</code></td>
<td>Answers a <code>Natural</code> number equal to <code>anInteger</code>, which must be less than the base of natural numbers.</td>
</tr>
</tbody>
</table>

Private instance methods for class `Natural`:

<table>
<thead>
<tr>
<th>Method</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><code>modBase</code></td>
<td>Answers a small integer whose value is the receiver modulo the base of <code>Natural</code> numbers.</td>
</tr>
<tr>
<td><code>divBase</code></td>
<td>Answers a <code>Natural</code> whose value is the receiver divided by the base of <code>Natural</code> numbers.</td>
</tr>
<tr>
<td><code>timesBase</code></td>
<td>Answers a <code>Natural</code> whose value is the receiver multiplied by the base of <code>Natural</code> numbers.</td>
</tr>
<tr>
<td><code>plus:carry: aNatural c</code></td>
<td>Answer the sum $\text{self} + \text{aNatural} + c$, where $c$ is a carry bit (either 0 or 1).</td>
</tr>
<tr>
<td><code>minus:borrow: aNatural b</code></td>
<td>Provided the difference is not negative, answer the difference $\text{self} - (\text{aNatural} + b)$, where $b$ is a carry bit (either 0 or 1). If the difference is negative, halt the program with a checked run-time error.</td>
</tr>
<tr>
<td><code>divmod:with: v aBlock</code></td>
<td>Computes quotient $q$ and remainder $r$ when $\text{self}$ is divided by divisor $v$, which should be a small integer. Answers the result of sending <code>(value aBlock q r)</code>.</td>
</tr>
<tr>
<td><code>compare:withLt:withEq:withGt: aNatural ltBlock eqBlock gtBlock</code></td>
<td>Compares <code>self</code> with <code>aNatural</code>. If <code>self</code> is smaller than <code>aNatural</code> evaluate <code>ltBlock</code>. If they are equal, evaluate <code>eqBlock</code>. If <code>self</code> is greater, evaluate <code>gtBlock</code>.</td>
</tr>
</tbody>
</table>

Figure 1: Suggested private methods for class `Natural`, list representation
first:rest:, which I recommend in fig. 1, very carefully. If both its arguments are zero, first:rest: must return an empty natural number. If either argument is nonzero, it must return a nonempty natural number.

3.1 Notes on private methods

Figure 1 lists private methods that I recommend for a list-based implementation of class Natural. (My recommendations for an array-based implementation are in the book.) Here are some notes:

- Private methods modBase, divBase, and timesBase are all simple manipulations of lists of digits. None of them should require any arithmetic, and only timesBase might require allocation.

- Division and modulus are always more easily done together than separately. Allocating a pair to hold quotient and remainder is a viable solution, but I recommend instead to use private method divmod:with:, which uses continuation-passing style. I have tried it both ways, and the continuation-passing code is significantly cleaner.

- The methods = and < which are required by class Magnitude are relatively easy to implement, but the obvious version of < also invokes = recursively, making the whole computation quadratic. To keep the cost of comparison linear, I instead recommend a trivalent comparison method compare:withLt:withEq:withGt:. Its interface is modeled on the ifTrue:ifFalse: method from class Boolean.

To help with debugging, I define other private methods:

- I define a printrep method, which prints the digits of a natural number, separated by slashes.

- I also define a do: method, which iterates over the digits of a natural number.

- I define a rep method that answers a list of digits, which I can use in unit tests. (I did my initial testing on base 16. Once everything was working, I shifted to a much larger base.)

- To simplify unit testing, I define a compare: method, which uses compare:withLt:withEq:withGt: to answer a symbol.

You may not need all these debugging methods—for me, the most useful was printrep.
3.2 What methods to define where

I'm recommending two subclasses of class `Natural`, and I believe that each of the private methods should be implemented separately on each of the subclasses. By contrast, methods `+`, `-`, `div`, `mod`, `=`, `<`, `decimal`, and `print` can all be defined just once, on class `Natural`. (I found no benefit to defining `*` on class `Natural`; it was easier just to do the two subclasses.)

3.3 Other hints

I recommend beginning by defining the methods on empty natural numbers—the ones that represent zero. Most implementations are simple indeed; for example, zero times anything is zero. Addition and subtraction require a little care; for example, you can sometimes add a number to zero by exploiting the equation $0 + n + c = n + 0 + c$, but this equation is useful only if $n \neq 0$.

Implementing the proper API for subtraction is one of the more annoying bits, because if the difference would be negative, you have to invoke an error continuation. My ML code uses exceptions, and exceptions make error-detection easy—but Smalltalk doesn’t have them. You could emulate exception-based error detection in Smalltalk, but emulating exceptions requires a gnarly tangle of continuations, and I recommend against it. The easy way out is to resign yourself to making two passes over the digits, and just compare the minuend with the subtrahend. Don’t try subtracting with `minus:borrow`: unless you know the difference is nonnegative.

If you’re trying to emulate the ML versions of the arithmetic operators, you might think that to do the comparable case analysis, you would need double dispatch. This much is true: if you wanted to do the same case analysis, you would have to use double dispatch. But who wants to do case analysis? Case analysis is the enemy! We never want to do case analysis. The beauty of the design I have sketched is that we can accomplish all the case analysis we need simply by dispatching to methods defined on the two subclasses of class `Natural`; no double dispatch is required. Try to understand how it works—this would make a good exam question.

3.4 What order to tackle the methods in

If you’re using lists, order is not as crucial as for arrays, but here’s the order in which I might tackle things:
1. Start with the class methods, including the private ones, and the private initialization methods. Don’t overlook public class method `new`:.

2. Next, all the methods on the empty natural number (zero).

3. Next, the methods defined on class `Natural`, like `decimal`, `+`, `print`, and so on. (For detailed advice on `decimal` and `print`, see page 12 of this handout.) I would write these methods first even though you won’t be able to test them: writing these methods will help you understand the APIs to the private methods.

   The methods on nonempty (nonzero) natural numbers should be broken down a little more finely.

4. Methods `isZero`, `modBase`, `divBase`, and `timesBase` are good starting points.

5. Next `printrep`.

6. Next `plus:carry:`, to get your feet wet with a binary arithmetic operation.

7. Next `divmod:with:`, so that you can try out `decimal` and make everything easier to debug.


9. Next `*`, which requires a bit more attention to detail than the others.

10. At the finish, I recommend `compare:withLt:withEq:withGt:`, which is mostly about passing and allocating continuations—just like the Boolean-formula solver.

### 3.5 A word about unit testing

If you define `rep`, you can unit test most of the private methods as you go along. Once you have the `decimal` method implemented, you can unit test the private methods with a little more confident. Some of the public methods can be tested as you go, and some (especially `−`), will be harder to test until almost everything is working.
4  Details of class Natural, array version

This section suggests some implementation details suitable for class Natural, assuming that you are representing a natural number as an array of digits. I suggest you implement class Natural in three stages, testing extensively at the end of each stage.

Stage I — Basics

1. Recommended private methods digit:, digit:put:, and makeEmpty: manipulate the array of digits that represent the number. Implement these methods first.
   Remember that digit: should work with any nonnegative argument, no matter how large.

2. Next, method doDigitIndices:. This method has to do with indexes, not with digits themselves. This way it can be used for mutation.

3. Once you have access to the digits, you can define trim, which removes unneeded leading zeroes. This method is meant to mutate the receiver.

4. Once trim is written, you can write digits:, to initialize a newly allocated bignum.

5. Now you can define the class method new:, which is your first public method—it creates a new object of class Natural.

Stage II — Simple functions

6. At this stage, a temporary print method is probably next—it will help you debug. For now, just print your representation.

7. Method isZero should be straightforward at this point. You’ll need a loop.

8. With access to the digits, you can write =. You can probably exploit private methods digit: and doDigitIndices: to make comparison relatively easy. You will need to be careful when comparing bignums of different degree, but there is a simple, elegant way to compare numbers of different degrees—try to find it.
Stage III — Arithmetic

9. The heart of your arithmetic implementation will be the two methods \texttt{set:plus:} and \texttt{set:minus:}. They depend on the digit methods above. A loop driven by \texttt{doDigitIndices:} may be helpful.

10. You can now implement addition with method \texttt{+}. Use \texttt{set:plus:}, and possibly also \texttt{trim} or \texttt{makeEmpty:}.

11. Subtraction is more complicated because it can fail: the difference of two natural numbers is not always a natural number. But the code should otherwise be similar to the addition case. In particular, building \texttt{subtract:withDifference:ifNegative:} on top of \texttt{set:minus:} should be analogous to your construction of \texttt{+}.

12. With natural-number subtraction in hand, you can now implement the standard methods \texttt{-} and \texttt{<}. You should be able to get everything you need from \texttt{subtract:withDifference:ifNegative:}, without having to use lower-level methods of class \texttt{Natural}.

13. To implement short division, you work down from most-significant digit to least-significant digit. I recommend defining a private method \texttt{setSdiv:remainder} which is sent to an object of class \texttt{Natural}, along with a one-digit divisor of class \texttt{SmallInteger}. The method mutates the receiver, dividing it by the divisor, and answering the remainder, also of class \texttt{SmallInteger}. It works by keeping a “current remainder” at each step. The current remainder is multiplied by base \(b\), added to the current digit, and the sum divided by the divisor. The quotient becomes a part of the result, and the remainder goes into the next step. The final remainder is what is answered from the method.

With \texttt{setSdiv:remainder} working, you can then implement the two public methods of short division:

- Method \texttt{div:} makes a copy of \texttt{self}, sends \(\texttt{(setSdiv:remainder copy divisor)}\), and answers \texttt{copy}.
- Method \texttt{mod:} makes a copy of \texttt{self}, then answers the result of sending \(\texttt{(setSdiv:remainder copy divisor)}\).

14. Multiplication is the most complicated operation of all. You will want to allocate a new number with \texttt{makeEmpty:} and initialize it to zero. Then, as suggested in the book, you’ll need a double sum to add in all the partial products. A doubly nested \texttt{doDigitIndices:} loop
will help. To manipulate the partial products, methods `digit: and digit:put:` are essential. Finally, use `trim` to control the growth of your bignums.

**Stage IV — Decimal conversion and printing**

15. Method `decimal` must answer a standard `List`. To convert natural number `n` to a list of decimal digits, I recommend initializing an empty list, then keep following these steps:

- As long as `n > 0`, use `addFirst:` to add `n mod 10` to the front of the list of digits, and replace `n` by `n div 10`.
- If `n = 0`, you’re almost finished. Just make sure the list of digits isn’t empty—if it is, add zero to it.

This algorithm can work with any representation using just `div:` and `mod:`. If you are using arrays, you can also use `setSdiv:remainder` directly—if you are careful.

16. Once you have `decimal`, `print` is easy:

```
(method print ()
   (do: (decimal self) (block (x) (print x))))
```

Debugging just got easier.

5 Details of large integers

The book defines class `LargeInteger`, but this definition is good enough only for homogeneous arithmetic on large integers, not for mixed arithmetic on large and small integers. You will need to add methods that add to, multiply by, or compare with a small integer. Here’s one example:

```
(method smallIntegerGreaterThan: (anInteger)
   (> self (asLargeInteger anInteger)))
```

You’ll need similar methods for addition and multiplication.

For testing, include this `decimal` method in class `LargeInteger`:

```
(method decimal () (locals decimals)
   (set decimals (decimal magnitude))
   (ifTrue: (negative self)
      {(addFirst: decimals '-)})
   decimals)
```

12
You will need to have implemented the `decimal` method on class `Natural`.

Once you’ve gotten this far, `LargePositiveInteger` and `LargeNegativeInteger` will be relatively straightforward. The list of methods and hints given in the book should get you through. You will lean heavily on your `Natural` methods, but only the `public` methods. These are the methods of class `Magnitude`, together with the methods listed in Figure 10.17 on page 842.

Here are a few example methods of class `LargePositiveInteger` from my solution:

```plaintext
(method negative () false)
(method strictlyPositive () (not (isZero self)))
(method + (anInteger) (addLargePositiveIntegerTo: anInteger self))
(method addLargePositiveIntegerTo: (anInteger)
  (withMagnitude: LargePositiveInteger (+ magnitude (magnitude anInteger))))
```

You’ll need a complete set of methods `negated`, `print`, `negative`, `nonnegative`, `strictlyPositive`, `+`, `*`, `addLargePositiveIntegerTo:`, `addLargeNegativeIntegerTo:`, `multiplyByLargePositiveInteger:`, and `multiplyByLargeNegativeInteger:`. (You’ll also need a `div:` method, but it can send `error`.) This design reuses the `LargeInteger` methods as much as possible.

### 6 Details of small integers with overflow detection

Getting mixed arithmetic to work requires a major overhaul of the `SmallInteger` class. Here are some illustrative methods:

```plaintext
(class SmallInteger SmallInteger ; overwrite SmallInteger with new class
 ()
 (method asLargeInteger () (new: LargeInteger self))

 (method + (aNumber) (addSmallIntegerTo: aNumber self))
 (method addSmallIntegerTo: (anInteger)
   (value (addSmall:withOverflow: self anInteger
     {(+ (asLargeInteger self) anInteger)})
   )
 (method addSmall:withOverflow: primitive add:withOverflow:)
 ...
```

The coercion method `asLargeInteger` enables mixed arithmetic. The three addition methods enable both mixed arithmetic (via double dispatch) and overflow detection (via primitive method, when adding two small integers).

1. You will need to replicate the addition structure for multiplication.
2. You will need to replace the primitive subtraction method with the classic “subtract from me by adding a negated argument.”

3. You will need to implement negated using a primitive method that can detect overflow.

4. To support mixed arithmetic, you will have to implement all the methods that get dispatched when + or * is sent to a large integer: addLargeNegativeIntegerTo:, addLargePositiveIntegerTo:, multiplyByLargeNegativeInteger:, and multiplyByLargePositiveInteger:.

5. You will have to use double dispatch to implement <, and you will also have to replace the primitive >.

6. You will need to reimplement the = method, probably by subtracting and comparing the difference with zero. You would benefit from implementing private method isZero as well.